Thank you for this opportunity to testify on the secondary proposals pending before the State Board. In previous testimony (9/17/08) Education Law Center raised concerns that major changes were needed if the proposals were to succeed in improving outcomes and achievement for all students. These changes included:

- A clear proposal for how new state assessments would be used to determine eligibility for high school graduation.
- An implementation plan for how schools & students would meet the new standards, particularly those not meeting current ones
- A cost/capacity study identifying what successful implementation will require at the state & district levels
- More flexibility for the development of multiple pathways to help all students to meet NJ graduation standards
- A research/evaluation design to track the impact of the proposals

In our view, none of these issues has yet been satisfactorily addressed. Accordingly, we urge the Board to take several steps that would, at least in part, address our most critical concern: namely, that these proposals could have a significant negative impact on NJ high school graduation and dropout rates.* To avoid this outcome, which no one wants, we propose the following:

First, we urge the Board to add an "opportunity to learn" provision clearly stipulating that no new high stakes exams will be mandated for high school graduation until the Department can

---

* For a summary of relevant research on this issue, see the literature review attached to ELC's testimony.

certify that all students have access to the courses and programs required to adequately prepare for them.

Second, we urge the Board to direct the Department to conduct a full evaluation of the Phase I mandates imposed this year before mandating Phases II & III.

We believe there are compelling reasons to include these provisions in the secondary proposals. Indeed, explicit commitments to equal opportunities to learn and the transparent evaluation of the impact of Board policies are basic prerequisites for fair and effective reform of the State assessment system.

The Department's most recent survey of graduation requirements in NJ high schools (see table below) shows that currently only 55% require Biology, 51% require Alg. I, 43% Geometry, 37% Algebra II and 34% Chemistry. The survey clearly shows that a majority of school districts will need to raise local graduation standards to meet the new mandates. What the survey does not tell us is whether local districts have the staff, facilities and other resources needed to do this. It is incumbent on the Board and the Department to document this capacity before imposing more high stakes graduation tests on students, and to assure that accountability is imposed first on those responsible for providing equal educational opportunity, rather than on young people who do not yet have it.

The Department's plans for raising graduation standards have not been accompanied by any study of what it will take to reach them, particularly for schools & students not meeting current standards.

Indeed the economic impact statement accompanying the proposals continues to declare, "there is no reason to anticipate that such curricular modifications would involve increased expenditures for school districts." This assertion is neither credible nor convincing.

We have previously provided Board members with documents from states, like Connecticut, which have projected costs of hundreds of millions of dollars to implement similar reform initiatives (costs which have led Connecticut and other states like Oregon and Pennsylvania to defer these mandates in the current economic climate.)

However, the fairest, most reliable way to address these concerns is a transparent look at the real impact of the Department's proposals. Phase I of the proposed mandates is already part of the current SFRA regulations. They require all freshmen entering in Sept. 08 to complete Biology, Alg. I & 4 yrs of college prep English before graduation. The regulations also require all middle and high schools to implement "personalization measures" as part of the "secondary transformation" needed to support higher levels of academic success. The Phase II course requirements for Geometry, Chemistry & a third lab science, do not go into effect until the entering freshman class in Sept. 2010.

This provides a window of opportunity for the State Board to do some "data-driven decision-making" of its own by directing the Department to conduct a full public evaluation of the Phase I mandates before mandating Phases II & III. Such an evaluation should document implementation progress and challenges, including enrollment changes in the required courses, pass/fail rates, student scores on the pilot assessments now in development, supports for struggling students and special needs populations, and the need for additional staff, facilities, professional development. Such an evaluation would go a

long way toward replacing an abstract discussion about promoting "higher expectations" for all students with some hard information about what it takes to reach them. It would also give the Dept. more time to build the broad consensus that clearly does not now exist to support an urgently needed secondary reform initiative.

The proposed mandates will affect every student, educator, school & community in the state for years to come. For these reasons, we strongly urge the Board to include an "opportunity to learn" guarantee and a Phase I evaluation plan before endorsing the Department's proposals.

Current High School Graduation Requirements By District Factor Group

Required	Abbott	A	B	CD	DE	FG	GH	I	J	Total
Algebra I	83%	73%	71%	46%	53%	44%	34%	40%	11%	51%
Geometry	77%	65%	66%	31%	35%	36%	32%	35%	11%	43%
Algebra II	71%	57%	57%	27%	27%	25%	32%	33%	11%	37%
Biology	75%	65%	69%	58%	53%	42%	36%	58%	67%	55%
Chemistry	54%	35%	46%	19%	27%	25%	23%	42%	56%	34%
No. of schools	*48*	*37*	*35*	*26*	*51*	*36*	*47*	*43*	*9*	*318*

Source: NJDOE 2008 Graduation Survey

Attached:

Summary Literature Review on the Impact of High Stakes Testing on Secondary Schools and Students
Prepared by Dr. Michelle Fine, CUNY Graduate Center
Appendix C: NJ's Special Review Assessment: Loophole or Lifeline?

Appendix C:

Summary Literature Review on the Impact of High Stakes Testing on Secondary Schools and Students

Since the emergence of public schools, the use of high stakes testing as a policy tool has been fiercely debated, with support for and criticism of standardized testing articulated by varied interest groups (Kliebard, 1991). Testing advocates using a humanist perspective argued that standardized exams measured students' knowledge about classic concepts they ought to master. Developmentalists promoted the use of testing to track students' individual propensities and interests. Social efficiency educators encouraged testing, hoping it would reduce the costs of public education. However, not all educators agreed with the premise of testing. George S. Counts of the University of Chicago wrote extensively about the deplorable classification of students and the reduction of curricula into minute units of work provoked by testing (Kleibard, p. 162). In 1930, he argued that the "feverish and uncritical fashioning of tests in terms of the existing curriculum and in the name of efficiency has undoubtedly served to fasten upon schools an archaic program of instruction and a false theory of the nature of learning" (as quoted by Kliebard, p. 162).

As testing became more commonly practiced in schools, the purposes of the examination began to swell — from recognizing students' needs and strengths to signifying a uniform curriculum, justifying stratified or tracked curricula, validating the worth of a high school diploma, rationalizing admissions or graduation, solidifying a "harmonious" system of schooling, aligning skills with the work place or higher education, monitoring teacher quality, punishing schools and educators, and now defending (or undermining) the high school institution to taxpayers. When the Committee of Ten came together in 1893, Resnick (1985) tells us there was no agreed upon understanding of what should be taught, when, how or to whom. It was difficult to determine the significance of the high school. In this political and educational context, the impetus for the original exit exam developed.

Within the debate about the efficacy of the SRA, one can hear traces of these historic arguments about the use of testing to clarify unresolved schooling issues, including the expectations for graduates and the value of the diploma.

Current Literature

In public debate today, testing has been argued to identify talent, diagnose problems, motivate performance, end social promotion, raise academic standards, increase the quality of schools, undermine public education, reduce racial and ethnic gaps, improve our international competitive edge and measure achievement (Greene et al., 2004). But high stakes testing has also been widely criticized for narrowing curricula and instructional quality, diluting classroom learning, increasing drop outs and push outs, reducing graduation rates, encouraging cheating by educators and students, and penalizing students and teachers for the educational system's inequalities (see Heubert and Hauser, 1999; Linn 2000).

Across centuries and places, it seems clear that politics and education have been intensely inter-related. Nowhere is this braiding more evident than in the current "accountability" debates. As Dan Korentz has argued, as the political significance of test scores heightens, the likelihood of distortions, misuse and cheating rises in kind (1992). We find ourselves today in the heat of highly contentious debates about testing and public education. Below we review some of the controversies and the empirical evidence surrounding high stakes testing: its effects on teaching, learning, graduation and dropping out; questions of validity and reliability, and alternative accountability platforms.

Ninth grade retention, graduation and dropout rates: Who is most affected by high stakes examinations?

> "Perhaps the most adverse unintended consequence of NCLB is that it creates incentives for schools to rid themselves of students who are not doing well, producing higher scores at the expense of vulnerable students' education. Studies have found that sanctioning schools based on average student scores leads schools to retain students in grade so that grade level scores will look better (although those students ultimately do less well and drop out at higher rates), exclude low-scoring students from admissions and encourage such students to transfer or drop out." (Darling-Hammond, 2007, p. 16)

High stakes exit exams have been consistently associated with decreased graduation rates, heightened dropout rates and reports of ninth grade retention. Many studies, summarized below, document an inverse relationship between high school graduation rates and the introduction of single-high stakes exit exams. Most significantly, the negative impact of standardized testing bears disproportionately adverse consequence for low-income students, students of color, and immigrant students. The Center on Education Policy reports that in 2006, 58 percent of White students but 76 percent of the nation's minority high school students were enrolled in public schools in the 22 states with exit exams; by 2012, 63 percent of White students and 81 percent of "minority" (Latino, African American, Asian/Pacific Islander, American Indian/Native Alaskan) students will be required to pass these exams (Center on Education Policy, 2007a).

- Haney (2000) demonstrates that the Texas Assessment of Academic Skills (TAAS) has had a particularly adverse impact on African American and Mexican American students, causing increased attrition rates and increased retention in the ninth grade. Haney also provides evidence that the exclusion of students with disabilities and students who dropped out of school in Texas helped produce "gains" in test scores.
- Amrein and Berliner (2002) used an archival time series research design to look for changes in dropout rates, high school graduation rates and enrollment in GED programs after exit exams were introduced in 16 states. The researchers conclude that high school exit exams led to higher dropout rates, lower graduation rates, and increased enrollments in GED programs in the majority of states.
- Warren, Grodsky, Lee, and Kulick (2005) find that completion rates (however measured) are simply much lower in states with High School Exit Exam (HSEE) policies. For example, for the graduating class of 2000, the median state high school completion rate was 73 percent for states with no HSEE but only 61 percent in states with HSEE's.

- The Center for the Study of Testing, Evaluation and Educational Policy at Boston College has analyzed the negative consequences of Massachusetts' current graduation policy. According to Haney, et. al. (2007) Massachusetts bases graduation decisions on the Massachusetts Comprehensive Assessment System (MCAS) Math and ELA test results, both norm referenced tests, despite the fact that the original Education Reform Act of 1993 called for multiple measures and criterion referenced tests. Haney et. al. argue that "students supposedly 'failing' MCAS have been shown to score well on national tests" (Haney, 2007) and, even more problematic from the New Jersey perspective, the attrition rates between grades 9 and 10 have increased sharply since 1997, particularly for Black and Latino students and those in low income urban districts.
- Sandholtz et al. (2004) found that those districts in southern California that implemented high stakes exit exams to enhance student achievement and reduce achievement gaps took an "unanticipated turn, one that seemed to move away from the expressed purpose of setting uniform academic expectations and thus providing equality of educational opportunity" and instead created "standard gaps" which resulted in differentiated curriculum and instruction along "perceived students' academic abilities." (p. 1179)
- The Center on Education Policy (2007c) finds that "gaps persist in high school exit exam pass rates" with a "persistence of gaps in initial pass rates particularly among English language learners and students with disabilities." Students who are "free or reduce-priced lunch eligible" were consistently and significantly less likely to pass state exit exams than "all students."
- McDermott (2007) compared New Jersey, Massachusetts, Vermont, and Connecticut's accountability systems and found that across systems the introduction of single high stakes exit examinations tends to constrict academic opportunities and diminish outcomes for the very groups of students the policies intend to help. McDermott writes that policy makers and policy implementers "need to keep in mind that the schools currently most challenged by standards based reform implementation are the same schools that were not previously educating all students to high standards and that those schools serve the students who have been most harmed by racial and socio-economic inequality. To change schools, the state must not only direct its power against them, through sanctions, but also build the power of all educational institutions to do their work better. Otherwise, aspirations to expand the moral boundaries of educational governance will continue to degenerate into policies that appear to blame victims for their own injuries." (p. 111)

Validity and Reliability: What is being measured?

While most of the research on high stakes examinations has focused on unintended effects, a number of scholars have directed critical attention to the technical quality of high stakes testing, raising questions about validity and reliability. Some investigate the extent to which perceived gains on exit exams correlate with other measures of achievement. Others focus on controversies in scoring, cheating, interpretation and reporting of results.

- Klein et al. (2000) join Haney and colleagues when they raise serious questions about the validity of gains in TAAS scores. They compare TAAS scores to scores on the National Assessment of Educational Progress Test (NAEP), and assess achievement gaps between White students and students

of color on the TAAS and the NAEP. They have found that: "According to NAEP, the gap is large and increasing slightly and according to TAAS, the gap is much smaller and decreasing greatly" (p. 16). The authors argue that the large discrepancies between TAAS and NAEP scores raise serious validity concerns about the TAAS scores.

- In studies that attempt to determine the impact of High School Exit Exams (HSEEs) on actual student achievement (actual growth in subject matter knowledge) (Grodsky, Warren, and Kalogrides, 2005; Warren et al., 2005), research using NAEP achievement data and advanced statistical analyses have produced *no viable evidence* that, in the 1990s, exit exams have done anything to significantly improve student knowledge of reading, mathematics, or science. Previous research in this area (see Carnoy and Loeb, 2002) has relied on state NAEP data, which was available in only 41 states prior to 2003. Because state HSEE policies change frequently over time, the association between student achievement and state NAEP scores on the long-term is tenuous.
- Cronbach et al. (1997) advise that agencies responsible for educational assessments should "make clear how uncertainty is associated with any score or summary, particularly with any report released to the public...." (p. 373). Given that student performances are part of larger measures of classrooms and schools, these authors emphasize that a standard error (SE) rather than a reliability coefficient "should be used to describe the uncertainty associated with scores for either individual students or schools...and should take into account constant as well as variable errors associated with tasks and scorers." (p. 396)

Compliance Behaviors: How do schools respond?

Interestingly, both critics and supporters of high stakes testing agree that strong external mandates and sanctions encourage schools to comply with testing policies. These effects are even more pronounced in schools serving poor children.

A number of empirical studies have found that externally imposed high stakes testing mandates tend to increase compliance often at the expense of many important elements of teaching, including critical thinking and creativity. For example:

- Nichols and Berliner (2007) found numerous examples across the country in which teachers dedicate substantial time to test preparation. Gail et (1999) calculate that elementary teachers spent more than 20% of their time practicing for high stakes tests. Nichols and Berliner estimate that this is the equivalent of 36 days of test prep. Even more dismaying, the authors calculate that "28 percent of those teachers report spending more than 60 percent of their time practicing for state tests...or over 100 of 180 days of instruction spent in various forms of test preparation." (p. 123)
- Diamond and Spillane's (2004) study of Chicago elementary schools found that lower-performing schools' focus on sanctions led to cosmetic rather than real changes in classroom teaching, emphasized impressing external observers rather than improving instruction, and anchored efforts to improve the performance of certain students within "benchmark grades and in certain subject areas," rather than all students.

- In Texas, McNeil and Valenzuela (2000) found that the Texas Assessment of Academic Skills (TAAS) as a management system has encouraged teachers to aim school instruction at the lowest level of information and skills, at the expense of complex assignments and critical analysis. McNeil and Valenzuela found that teachers spend classroom time drilling on practice examination materials and reducing subjects to an artificial treatment of "isolated skills and fact fragments," which do not necessarily "enable students to use these components in other contexts." (p. 6)
- Comparing the classroom cultures of two reading lessons taught by the same teacher, Valli and Chambliss (2007) found that as the class came to be dominated by test preparation, students were far less likely to engage in dialogue and the teacher was far less likely to scaffold student learning by connecting class work to students' real-life experiences. Additionally, teachers were more likely to ask simple questions and students were more likely to respond with simple answers. (p. 16)
- The Center on Education Policy (2005) conducted case studies in both Virginia and Maryland and found that exit exams appear to be increasing anxiety for both teachers and students, as well as detracting from instructional time. Particularly in Virginia, teachers reported feeling stressed about losing their jobs and the overall impact of exams on their practice. Students described the school atmosphere as intense, referring to teacher and administrator stress as well as to the rapid pace and repetition of material coverage.
- The Center on Education Policy (2007) also published case studies on Austin, Texas and Jackson, Mississippi where teachers and students report, with enthusiasm, the frequent use of test-preparation strategies during class time. Data from interviews and surveys revealed that teachers are using common test-preparation strategies, including spending class time reviewing test-taking skills, using previous exit exam questions on classroom tests, and practicing sample test questions during class. In their press release, CEP researchers claim, "[A]ssessments appear to have increased instructional time in tested subject areas in both districts, often at the expense of other high school learning experiences and electives....Teachers at [schools serving greater numbers of lower-income and minority students] reported that the Texas Assessment of Knowledge and Skills [TAKS] had influenced every aspect of the curriculum, while their peers at a school serving greater numbers of higher-income and white students reported that the TAKS was an inconvenience." (CEP, 2007b, press release)
- Firestone and Mayrowet's (2000) classroom observations conducted in England, Wales, Maryland and Maine, suggest that testing and the associated external pressures may facilitate changes in the content taught, but not in instructional strategies. Firestone and Mayrowet claim that single high stakes examinations may "primarily promote short term accommodations but not deeper learning." (p. 20)
- Coleman (1998) argues that "Tests, in short, should be instruments used by educators to help students achieve their full potential...policy makers and the education community must work to guarantee that the establishment of high stakes standards for all students does not unfairly result in the denial of educational opportunity for any one student." (pp. 82-83) To support student learning and assessment, Coleman argues that schools should establish compensatory or tutorial supports to ensure that all students have the same basic and fair opportunity to master the material tested; provide multiple opportunities for test takers to take the test, and consider academic factors in addition to the tests scores that may affirm or challenge the high stakes conclusions derived from the test scores.

Accountability Loopholes

> "[P]olicies that reward or punish schools have created a distorted sense of accountability, encouraging manipulations of student placements (including pushing students out) as well as encouraging staff to opt for school placements where school stability is higher and students are perceived more likely to meet score requirements" (Darling-Hammond, 2004, p. 1).

Accountability policies anchored on single high stakes assessments often provoke the creation of "accountability loopholes." (Vazquez, 2006) Shriberg and Shriberg (2007) note that under current policy, many schools "rather than supporting students who are otherwise making progress toward graduation, but who are for various reasons (such as language barriers, lack of exposure to test content, past history of low test scores, and so on) unlikely to pass proficiency exams...do better by encouraging students to drop out or transfer so that their test scores do not trigger draconian penalties" (2007, p. 80).

- Booher-Jennings (2005) found teachers used data to facilitate what she calls "educational triage," diverting additional resources and attention to students believed to be on the threshold of passing the test, in order to create the impression of improvement.
- Valenzuela et al. (2006) found that high schools that increased retention or disappearance of students in 9th grade were able to increase their exit scores. The overall disappearance rate more than doubled for ELL (English Language Learning) students.
- Vazquez-Helig (2006) found that exit exams create incentives for schools and districts to utilize accountability loopholes, leading to an escalation of deleterious outcomes disproportionately impacting minority students including sharp increases in 9th grade student retention, associated with high levels of student dropout and disappearance. They also identify other practices such as disciplinary measures as a means to push out students, skipping students past key testing grades, and transferring students to non-traditional settings.
- Jacob and Levitt (2001) developed an algorithm for detecting teacher cheating on standardized tests. Using data from the Chicago Public Schools Iowa Test of Basic Schools (ITBS) scores, the authors estimate that serious cases of teacher or administrator cheating occur in four to five percent of elementary schools each year. Moreover, the authors conclude that "teacher cheating appears quite responsive to relatively minor changes in incentives." (p. 34) Their results suggest that introducing high stakes testing without appropriate safeguards against cheating would likely lead to widespread corruption.

Deep Accountability Systems: Multiple Pathways to Graduation

Accountability expert Linda Darling-Hammond has written extensively on the corrupting influences of high stakes single-test accountability practices as well as designs for more valid and equitable accountability policies. In states and school districts that have focused on broad and deep accountability policies that are both valid and equitable, accountability reforms tend to be embedded in the practices of teaching and learning; assessment strategies are built into professional development and capacity building of educators; students are provided extra supports as needed and opportunities for revision in order to demonstrate their mastery of state standards. That is, *states and districts with strong achievement*

and graduation rates typically have multiple assessment pathways by which students may satisfy graduation requirements.

The Center on Education Policy, in *State High School Exit Exams: A Challenging Year,* (August 2006) has profiled some of these innovative states committed to high rates of achievement and graduation through multiple pathways. A few examples may help readers imagine how to combine valid and equitable assessment strategies with rigorous state standards:

- In Connecticut, districts award diplomas based on a combination of local tests, student academic records, and state test results, but districts cannot deny a diploma solely on the basis of the state test. (Delaware, Maine, Rhode Island, and Wyoming have slight variations on this theme.) Students must pass some assessment requirements for a diploma, but there is flexibility in what these requirements are.
- In Oregon, students in the class of 2007 and beyond will receive a regular diploma only if they demonstrate proficiency using some locally developed assessment components, most of which are portfolio-based assessments.
- In Virginia, students can substitute scores from other tests such as Advancement Placement exams.
- In 2006, Utah decided not to withhold diplomas from students who failed the state competency test.
- Several other states leave it up to districts whether to make assessment results a factor in awarding diplomas or permit that as an option (i.e., Kentucky, New Hampshire, Kansas, Missouri and Wisconsin). New Hampshire places just one condition on districts' decisions: namely, that the state assessment cannot be the sole factor in awarding a diploma.
- Nebraska has expressly created *no* statewide assessments, and instead encouraged the development of rich local assessments. Individual districts have the authority to decide the extent to which these assessments are factored into awarding diplomas. Teachers collaborate to craft assessments measuring students' mastery, pilot these assessments in their own classrooms, and meet on an ongoing basis to fine-tune them. This local model supports teachers in their expanding roles as instructional leaders, is woven into the curriculum, and mirrors the state's educational goals to deepen students' mastery.
- The New York Performance Standards Consortium represents an impressive model of public schools dedicated to high standards, rigorous performance assessments and broad based accountability. The New York Performance Standards Consortium currently includes more than 20 public schools in New York City and State that have collaboratively generated and now implement a series of performance standards which students must complete in order to earn a high school diploma.

The history of the Consortium begins in 1995, when New York's former Commissioner of Education, Thomas Sobol, granted these schools a waiver from the state's Regents exams, supporting them in their commitment to develop and utilize a "transparent, externally-validated performance-based assessment system — reliable across schools — that would assure complex teaching and in-depth learning of the state standards and beyond." (Foote, 2007, p. 8) Today students in these schools must take and pass the Language Arts and Mathematics Regents examinations but beyond that the schools collaboratively generate inquiry based, student centered, in depth research projects for youth to pursue to satisfy their

graduation requirements. With a common metric implemented across schools, students craft intensive discipline-based and inter-disciplinary projects which are assessed by panels of educators, students and university professors.

With a rigorous set of performance tasks aligned with state standards and an extensive external review of the schools and student work by a panel of university based and secondary educators, the Performance Standards Consortium schools have a very strong track record of high graduation rates for students who might be considered "high risk." Indeed, Foote has recently undertaken a follow up study of Consortium graduates to determine how they fare in higher education. Able to track a sample of 666 graduates from fifteen schools, Foote reports in *Phi Delta Kappan:*

> "...77 percent of Consortium school graduates attended 4-year colleges, 19 percent attended 2-year colleges, and 4 percent attended vocational or technical programs. Further, after a year and a half of college, the average GPA for graduates in the sample was 2.6 out of 4.0, which is approximately a B–. For students attending 4-year colleges, the average GPA was 2.7. For students attending 2-year institutions, the average GPA was 2.2.
>
> Of those in the sample who entered college within one year of high school graduation, 78 percent overall enrolled for a second year. Of those attending 4-year colleges, 84 percent enrolled for a second year. Of those attending 2-year institutions, 59 percent enrolled for a second year." (p. 361)

The Consortium schools have designed performance assessments that are rigorous, embedded in the school curriculum and highly student centered. With this research, Foote has demonstrated, that this performance-based approach to teaching, learning and assessment has high predictive validity for post-secondary success. The results of the Consortium, like so many alternative assessment systems, assume that "real accountability is achieved when an assessment system demands excellence not only from students, teachers, and principals, but also from itself, with an oversight mechanism for external validation as well as ample evidence of student success beyond test scores...including proof of an assessment system's predictive validity with data correlating the passage of specific assessments with subsequent performance in school, college, or the work force." (p. 363)

New Jersey has long been in the forefront of achievement and graduation rates and can now be a model for valid and equitable alternative pathways to graduation. Policy makers invested in making policy, legislative, and programmatic changes necessary to help all students graduate from high school prepared for work and/or for further education can learn much from the significant research on the adverse consequences of high stakes testing and the well-documented alternatives to single high stakes exit examinations proliferating throughout the nation.